Multilingual Dynamic Topic Model

Zosa, Elaine

INCOMA
2019-09-04

Zosa , E & Granroth-Wilding , M 2019 , Multilingual Dynamic Topic Model . in G Angelova , R Mitkov , I Nikolova & I Temnikova (eds) , RANLP 2019 - Natural Language Processing a Deep Learning World : Proceedings . International conference Recent advances in natural language processing , INCOMA , Shoumen , pp. 1388-1396 , Recent Advances in Natural Language Processing , Varna , Bulgaria , 02/09/2019 . https://doi.org/10.26615/978-954-452-056-4_159

http://hdl.handle.net/10138/307837
https://doi.org/10.26615/978-954-452-056-4_159

publishedVersion

Multilingual Dynamic Topic Model

Elaine Zosa and **Mark Granroth-Wilding**
Department of Computer Science
University of Helsinki
Helsinki, Finland
firstname.lastname@helsinki.fi

1 Abstract

Dynamic topic models (DTMs) capture the evolution of topics and trends in time series data. Current DTMs are applicable only to monolingual datasets. In this paper we present the multilingual dynamic topic model (ML-DTM), a novel topic model that combines DTM with an existing multilingual topic modeling method to capture cross-lingual topics that evolve across time. We present results of this model on a parallel German-English corpus of news articles and a comparable corpus of Finnish and Swedish news articles. We demonstrate the capability of ML-DTM to track significant events related to a topic and show that it finds distinct topics and performs as well as existing multilingual topic models in aligning cross-lingual topics.

2 Introduction

Dynamic topic models (DTMs, Blei and Lafferty, 2006) capture themes or topics discussed in a set of time-stamped documents and how the words related to these topics change in prominence over time. Other topic models have been proposed that aim to model time series data (Wang and McCallum, 2006; Wei et al., 2007; Hall et al., 2008). These models can be used to explore historical document collections to study historical trends, language changes (Frermann and Lapata, 2016) and track the emergence and evolution of certain subjects (Hall et al., 2008; Yang et al., 2011).

With the internet becoming more multilingual it is increasingly important to build cross-lingual tools to bridge different linguistic groups online. Fortunately, large multilingual datasets such as Wikipedia, the Europarl parallel corpus (Koehn, 2005) and other datasets assembled from crawling the web (Van Gael and Zhu, 2007) are also becoming widely available to researchers. This has led to the development of several multilingual topic models to infer topics from multilingual datasets. Examples include the polylingual topic model (PLTM, Mimno et al., 2009), multilingual topic model for unaligned text (MuTo, Boyd-Graber and Blei, 2009), and JointLDA (Jagarlamudi and Daumé, 2010). What is currently lacking are topic models for multilingual time-stamped data that can model historical and linguistic changes in a specific context. Digitalization efforts in libraries and archives, such as the Europeana collections[1], have made available online historical document collections from different European countries. Collections such as these are valuable resources for comparing historical trends in different countries. However, scholars and other interested parties may not possess the linguistic skills necessary to explore such data and would benefit from tools to automatically discover connections across linguistic boundaries.

In this paper, we present the multilingual dynamic topic model (ML-DTM), a novel topic model that captures dynamic topics from broadly topically aligned multilingual datasets. We extend a DTM inference method by Bhadury et al. (2016) to train this model.

In the following sections, we give a broad review of related work, discuss existing *dynamic* and *multilingual* topic models in more detail, and then give a description of our proposed combined model. We then demonstrate usage of this model on a parallel dataset and a comparable dataset of news articles and present our results. We show that this novel topic model learns aligned bilingual topics as demonstrated by the cosine similarities of learned vector representations of named entities. Table 1 summarizes the notations used in this paper. Code is available at: https://github.com/ezosa/multilingual_dtm.

[1] https://www.europeana.eu

Proceedings of Recent Advances in Natural Language Processing, pages 1388–1396,
Varna, Bulgaria, Sep 2–4 2019.

Symbol	Description
α	parameter for θ
β	hyperparameter for ϕ
ψ	hyperparameter for θ
θ	distribution of topics over a document
ϕ	distribution of words over a topic
D	set of documents
W_d	words in document d
N_d	number of words in document d, or $\|W_d\|$
Z_d	topic assignments of words in document d
K	number of topics
T	number of time slices
L	number of languages in the dataset
V	words in a vocabulary for language

Table 1: Summary of notations.

3 Related Work

Topic models capture themes inherent in document collections through the co-occurence patterns of the words in documents. Latent Dirichlet Allocation (LDA, Blei et al., 2003) is a popular method for inferring these themes or topics. It is generative document model where a document is described by a mixture of different topics and each topic is a probability distribution over the words in the vocabulary. In a document collection we can only observe the *words* in a document. Therefore, training a model involves inferring these latent variables through approximate inference methods.

In the case of documents with timestamps covering some time interval, such as news articles, we might want to capture *dynamic* co-occurence patterns that evolve through time. Dynamic Topic Model (DTM, Blei and Lafferty, 2006) divides time into discrete slices and chains parameters from each slice in order to infer topics that are aligned across time. DTM gives us a set of topic-term distributions that evolve from one time slice to the next. There are also other topic models for time-series data such as the Continuous Dynamic Topic Model (cDTM, Wang et al., 2008), a version of DTM that does not explicitly discretize time intervals. Dynamic Mixture Model (DMM, Wei et al., 2007) captures the evolution of documents across time and Topics over Time (TOT, Wang and McCallum, 2006) is a method that models the prominence of topics over time.

A limitation of LDA, as well as these dynamic models, is that it is not applicable to multilingual data. LDA captures co-occurences of words in documents and words from different languages would rarely, if ever, occur in the same document regardless of their semantics, as demonstrated by experiments on the Europarl corpus (Jagarlamudi and Daumé, 2010; Boyd-Graber and Blei, 2009). Multilingual topic models are developed to capture cross-lingual topics from multilingual datasets.

Polylingual Topic Model (PLTM, Mimno et al., 2009) is a multilingual topic model that extends LDA for an aligned multilingual corpus. Instead of running topic inference on individual documents as in LDA, PLTM infers topics for *tuples* of documents, where each document in the tuple is in a different language. PLTM assumes that the documents of a tuple discuss the same subject broadly and therefore share the same document-topic distribution.

Other topic models for multilingual data include Multilingual Topic Model for Unaligned Text (MuTo, Boyd-Graber and Blei, 2009) and JointLDA (Jagarlamudi and Daumé, 2010). MuTo attempts to match words between languages in the corpus and samples topic assignments for these matchings. JointLDA is a multilingual model that does not require an aligned corpus but requires a bilingual dictionary and uses concepts, instead of words, to infer topics where concepts can be entries in the bilingual dictionary.

In this work we will focus on DTM and PLTM because we want to capture topic evolution in multilingual settings without using additional lexical resources such as dictionaries.

3.1 Dynamic Topic Model

LDA uses Dirichlet and multinomial distributions for inferring both topic-term distributions ϕ and document-topic distributions θ. The conjugacy of these distributions allow ϕ and θ to be integrated out leaving us only with the posterior distribution for topic-term assignments Z, which we can sample through Gibbs sampling (Griffiths and Steyvers, 2004). Inference in DTM, however, is

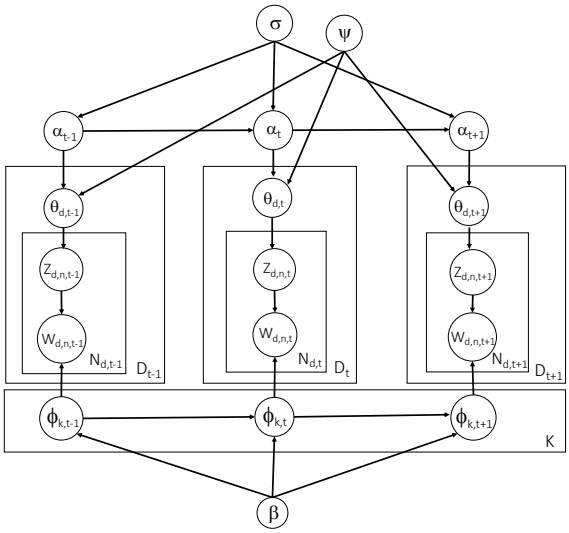

Figure 1: DTM for three time slices as shown in Bhadury et al. (2016).

more complicated due to the non-conjugacy of the distributions used in the model. Blei and Lafferty (2006) use variational Kalman filtering for topic inference, which does not scale well for a large number of topics and documents and large numbers of time slices (Bhadury et al., 2016; Wang et al., 2008). Bhadury et al. (2016) developed a method for inferring the posterior distributions of DTM with Gibbs sampling. In their method, the parameters α, θ, ϕ and Z are re-sampled during every iteration of the sampler.

The document-topic proportions θ, sampled for each document in each time slice, and the topic-term distributions ϕ, sampled for each topic in each time slice, are updated using Stochastic Gradient Langevin Dynamics (SGLD, Welling and Teh, 2011) which is based on Stochastic Gradient Descent (SGD). Figure 1 shows the plate diagram for DTM from Bhadury et al. (2016).

3.2 Polylingual Topic Model

The polylingual topic model (PLTM, Mimno et al., 2009) is an extension of LDA that infers topics from an aligned multilingual corpus composed of document tuples. Tuples are composed of documents in different languages that are thematically aligned, meaning that they discuss the subject in broadly similar ways. For instance, a news article in German and another article in English that report on the same event can compose a tuple.

Inference on PLTM can be done via Gibbs sampling where the topic assignment of each term $z^l_{d,n}$ is resampled during every iteration. Following Vulić et al. (2015), we provide the update formulae for the bilingual case for brevity. The update formulae for documents in languages x and y are:

$$P(z^x_{d,n} = k|z^x, z^y, w^x, w^y, \alpha, \beta) \propto \frac{m^x_{d,k} - 1 + m^y_{d,k} + \alpha}{\sum_{i=1}^{K} m^x_{d,i} - 1 + \sum_{i=1}^{K} m^y_{d,i} + K\alpha} \cdot \frac{v^x_{k,w_{d,n}} - 1 + \beta}{\sum_{i=1}^{|V^x|} v^x_{k,w_{d,i}} - 1 + |V^x|\beta} \quad (1)$$

$$P(z^y_{d,n} = k|z^y, z^x, w^y, w^x, \alpha, \beta) \propto \frac{m^y_{d,k} - 1 + m^x_{d,k} + \alpha}{\sum_{i=1}^{K} m^y_{d,i} - 1 + \sum_{i=1}^{K} m^x_{d,i} + K\alpha} \cdot \frac{v^y_{k,w_{d,n}} - 1 + \beta}{\sum_{i=1}^{|V^y|} v^y_{k,w_{d,i}} - 1 + |V^y|\beta} \quad (2)$$

where $m^x_{d,k}$ is the number of times topic k has been assigned to a word in document d written in language x and $v^x_{k,w_{d,n}}$ is the number of times word $w_{d,n}$, that is, the word at position n in document d, has been assigned to topic k. $|V^x|$ is the vocabulary size of language x. The first part of these formulae links the two languages together and is language-independent while the second part is language-specific.

Figure 2 shows the graphical representation of PLTM for l languages.

4 Multilingual Dynamic Topic Model

Here we combine the above *dynamic* and *polylingual* models to produce a *Multilingual Dynamic Topic Model* (ML-DTM). Figure 3 shows the diagram of ML-DTM for two languages and three time slices. Although we show only the bilingual case here for brevity, the model is applicable for any number of languages.

The inference method of Bhadury et al. (2016) was originally motivated by the need to speed up DTM inference for very large datsets. We apply it here to the combined ML-DTM model. We propose the following posterior conditional distribution for $\theta_{x,t}$ where x is a tuple index in the dataset:

$$p(\theta_{x,t}|\alpha_t, Z_{x,t}) \propto \mathcal{N}(\theta_{x,t}|\alpha_t, \psi^2 I) \times \prod_{l=1}^{L} \prod_{n=1}^{N_{d_l,t}} Mult(Z_{d_l,n,t}|\pi(\theta_{x,t})) \quad (3)$$

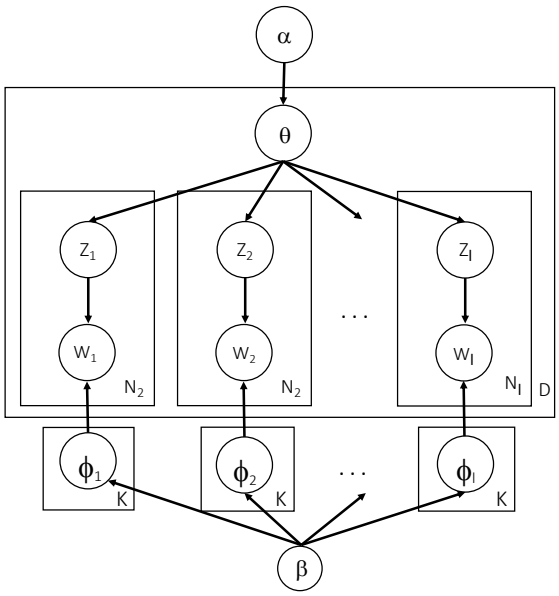

Figure 2: Polylingual topic model for l languages of Mimno et al. (2009).

Following Bhadury et al. (2016), the update equation to evaluate the gradient of $\theta^k_{x,t}$ becomes:

$$\nabla_{\theta^k_{x,t}} \log p(\theta_{x,t}|\alpha_t, Z_{x,t}) = \frac{-1}{\psi^2}(\theta^k_{x,t} - \alpha^k_t) + \sum_{l=1}^{L} C^k_{d_l,t} - \left(N_{d_l,t} \times \frac{\exp(\theta^k_{x,t})}{\sum_j \exp(\theta^j_{x,t})} \right) \quad (4)$$

where $Z_{x,t}$ are the topic assignments for the words in the documents in tuple x at time slice t; $C^k_{d_l,t}$ is the number of times topic k has been assigned to a word in document d_l at time t; and $N_{d_l,t}$ is the length of document d_l at time t.

Instead of evaluating $\theta_{d,t}$ for a single document as in monolingual DTM, we compute $\theta_{x,t}$ for a document *tuple*. The second term in (4) links the languages together by summing up the counts of each document in the tuple.

The equation for evaluating the gradient of the topic-term distributions $\phi_{k,t}$ is the same as in the original paper except that we compute separate distributions for each language since every language has a different vocabulary. This means that for each time slice, instead of updating K different ϕs (one for each topic), we will need to update $K \cdot L$ ϕs. Table 2 shows the dimensions of the parameters to be estimated.

Finally, the topic assignment $Z_{d_l,n,t}$ is sampled

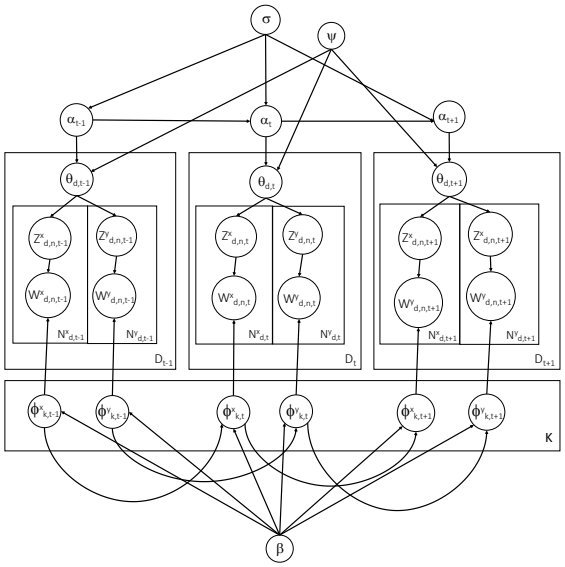

Figure 3: ML-DTM for two languages and three time slices.

Parameter	Dimension
α	$K \times T$
θ	$D^t \times K \times T$
ϕ	$\lvert V^l \rvert \times L \times K \times T$

Table 2: Dimensions of the sampled parameters in the multilingual dynamic topic model (ML-DTM). D^t is the number of document tuples in a dataset.

as in the original paper:

$$P(Z_{d_l,n,t} = k|\theta_{x,t}, \phi^{w_l}_{k,t}) \propto exp(\theta^k_{x,t}) exp(\phi^{w_l}_{k,t}) \quad (5)$$

where w_l is a word from the vocabulary of language l.

5 Evaluation

5.1 Datasets

We ran experiments on ML-DTM with two kinds of data: a parallel dataset and a thematically-comparable one.

The DE-NEWS parallel dataset consists of German news articles from August 1996 to January 2000 with English translations done by human volunteers[2]. This dataset covers 42 months with an average of 200 articles per month. Since this is a parallel corpus there is no need to align the articles.

[2] http://homepages.inf.ed.ac.uk/pkoehn/publications/de-news/

For the comparable dataset, we use the YLE news dataset which consists of Finnish and Swedish articles from the Finnish broadcaster YLE, covering news in Finland from January 2012 to December 2018[3]. The Finnish and Swedish articles are written separately and are not direct translations of each other. We use existing methods for aligning comparable news articles (Utiyama and Isahara, 2003; Vu et al., 2009). Specifically, we create an aligned corpus by pairing a Finnish article with a Swedish article published within a two-day window and sharing three or more named entities. We want to have a one-to-one alignment in our dataset such that no article is duplicated, so we pair a Finnish article with the first Swedish article encountered in the dataset that fits the above criteria and remove the paired articles from the unaligned dataset. The unaligned dataset has a total of 604,297 Finnish articles and 228,473 Swedish articles and the final aligned dataset consists of 123,818 articles covering 84 months. A script for aligning articles using the method described is provided in the Github project associated with this work.

We tokenized, lemmatized (using WordNetLemmatizer for German and English and LAS (Mäkelä, 2016) for Finnish and Swedish) and removed stopwords for these two datasets and then used the 5,000 most frequent words of each language as the vocabulary for that language.

5.2 Cross-Lingual Alignment

We compare the cross-lingual alignment of topics of ML-DTM and PLTM by evaluating the similarity of the learned vector representations of named entities (NEs) that appear in both languages of the same dataset. This method is suggested by Vulić et al. (2015) on the basis that NEs tend to be spelled in the same way in different languages and can be expected to have a similar association with topics across languages. The K-dimensional vector of a NE w for language s is thus:

$$vec(w_s) = [P(z_1|w_s), P(z_2|w_s), ..., P(z_K|w_s)] \quad (6)$$

Under an assumption of a uniform prior over topics, this vector can be computed as:

$$P(z_k|w_s) \propto \frac{P(w_s|z_k)}{P(w_s)} = \frac{\phi_{l,z_k,w_s}}{Norm_{\phi_{s,.,w_s}}} \quad (7)$$

$$Norm_{\phi_{s,.,w_s}} = \sum_{k=1}^{K} \phi_{s,z_k,w_s} \quad (8)$$

$$vec(w_s) = \frac{[\phi_{l,z_1,w_s}, \phi_{l,z_2,w_s}, ..., \phi_{l,z_K,w_s}]}{Norm_{\phi_{s,.,w_s}}} \quad (9)$$

We then take the cosine similarities between the L different vector representations of the NE (for both datasets, $L = 2$).

We evaluate the cosine similarities of NEs that occur five or more times in each time slice. To make the comparison between PLTM and ML-DTM, we train one ML-DTM model on three time slices for 10 topics and three separate PLTM models for each time slice, also capturing 10 topics. We set $\alpha = 1.0$ and $\beta = 0.08$ for PLTM and $\alpha = 0.5$ and $\beta = 0.5$ for ML-DTM for both datasets, which achieved the best results of a small range of values tried. We did not, for now, perform more extensive optimisation of hyperparameters.

5.3 Topic Diversity

We also measure the *diversity* of the topics ML-DTM finds by computing the Jensen-Shannon (JS) divergence of every topic pair for each time slice for each language and averaging the divergences. Wang and McCallum (2006) used this method, though with KL divergence. It is desirable for the model to find topics that are as distinct as possible from each other.

We compare the diversity of the topics found by ML-DTM, trained as in the previous section, with the topics found by DTM. To make this comparison we train separate DTM models for each language in our two datasets, giving us four different models and compare the divergences of the topics found by these models with their ML-DTM counterparts. We use the Gensim implementation of DTM[4] where we set the chain variance to 0.1 and leave other parameters to be inferred during training. We train both ML-DTM and DTM on 10 time slices for 10 topics.

[3] https://www.kielipankki.fi/corpora/

[4] https://radimrehurek.com/gensim/models/ldaseqmodel.html

Time slice	# of NEs	PLTM	ML-DTM
Aug 1996	53	**0.880**	0.692
Sept 1996	65	0.876	**0.908**
Oct 1996	64	0.840	**0.885**

Table 3: Average cosine similarity of topic vectors for NEs over three time slices in DE-NEWS.

Time slice	# of NEs	PLTM	ML-DTM
Jan 2012	79	0.800	**0.896**
Feb 2012	71	**0.810**	0.796
Mar 2012	72	0.722	**0.745**

Table 4: Average cosine similarity of the vectors of NEs for three time slices in the YLE dataset.

6 Results and Discussion

Tables 3 and 4 show the average cosine similarity between NEs for each language in the DE-NEWS and YLE datasets, respectively. In the DE-NEWS data (Table 3), PLTM outperforms ML-DTM in the first time slice but ML-DTM performs better on the succeeding time slices. This is an encouraging result, considering that the parameters of ML-DTM at time slice t are estimated from adjacent time slices, adding a large degree of complexity to the model, whereas PLTM estimates parameters based on the current time slice only (PLTM has no concept of time).

For the YLE dataset (Table 4), ML-DTM shows an improvement in the first time and third slices and comparable performance in the second. The comparable nature of this dataset makes aligning NEs a more challenging task for both models. One way to improve performance on this task might be to use stricter criteria in aligning the dataset, such as pairing articles only if they were published on the same day or if they share more named entities.

We compare topic diversity of the topics found by DTM and ML-DTM. Tables 5 and 6 show the average JS divergence of every topic pair for five time slices in the DE-NEWS and YLE datasets, respectively. ML-DTM consistently learns more diverse topics than DTM for both datasets.

In Figure 4, we show the evolution of one topic found by ML-DTM trained on DE-NEWS. We show the top words of a topic about labor unions for the first eight months of the dataset. The English and German words are not exact translations of each other but we see similar or related words

Time slice	DTM English	ML-DTM English
Aug 1996	0.372	**0.655**
Sep 1996	0.368	**0.660**
Oct 1996	0.366	**0.657**
Nov 1996	0.365	**0.664**
Dec 1996	0.363	**0.650**
	DTM German	ML-DTM German
Aug 1996	0.315	**0.661**
Sep 1996	0.312	**0.670**
Oct 1996	0.310	**0.665**
Nov 1996	0.308	**0.638**
Dec 1996	0.306	**0.666**

Table 5: Topic diversity comparison between DTM and ML-DTM: average JS divergences of each topic pair for five months of the DE-NEWS dataset for English and German.

and NEs in each time slice. For instance, in August 1996 'employer' and 'arbeitgeber' both appear, as does 'einzelhandel' and 'retail'. In Sept 1996, 'kohl' is the top term for both languages (referring to former German chancellor Helmut Kohl). There are cases where German terms have no direct translation in English but an equivalent concept appears in the English topic. This is the case with 'lohnfortzahlung' (sick-leave pay) where the terms 'sick' and 'pay' appear on the English side; and 'steuerreform' (tax reform) where 'reform' appears on the English side as well.

A named entity, 'thyssen', appears in March 1997 in both languages but not in other months. This is because of an event that happened around mid-March where the German steel company Thyssen was being bought by competitor Krupp-Hoesch (also a top term in the German topic) prompting concerns about job losses[5].

Figure 5 shows the first six months of a topic about political news from the YLE dataset. The first two months has terms related to presidential elections. This refers to the Finnish presidential election in 2012, where rounds of voting took place in January and February 2012[6]. These time slices also mention the two candidates in the runoff election, Sauli Niinistö and

[5] https://www.nytimes.com/1997/03/19/business/krupp-hoesch-confirms-bid-of-8-billion-for-thyssen.html

[6] https://en.wikipedia.org/wiki/2012_Finnish_presidential_election

Time slice	DTM Finnish	ML-DTM Finnish
Jan 2012	0.332	**0.445**
Feb 2012	0.324	**0.465**
Mar 2012	0.322	**0.470**
Apr 2012	0.353	**0.498**
May 2012	0.357	**0.495**
	DTM Swedish	ML-DTM Swedish
Jan 2012	0.365	**0.480**
Feb 2012	0.360	**0.491**
Mar 2012	0.354	**0.497**
Apr 2012	0.388	**0.535**
May 2012	0.393	**0.537**

Table 6: Topic diversity comparison between DTM and ML-DTM: average JS divergences of each topic pair for five months of the YLE dataset for Finnish and Swedish.

Pekka Haavisto. Sauli Niinistö eventually won the election which explains why the next time slices ceases to mention Pekka Haavisto while 'niinistö' is still a prominent term. After March 2012, the topic stops talking about presidential elections and moves on to other political news. This gives us an insight into how the model can track significant events, such as high-profile elections, related to a topic. Another example is May 2012, where Greece ('kreikka' in Finnish, 'grekland' in Swedish) suddenly becomes a prominent term for both languages due to the Greek legislative elections which took place on 6 May 2012. The term 'syyria'/'syrien' appears in May and June, corresponding to the beginning of the Syrian Civil War.

Figure 6 shows the posterior probabilities of some terms related to the presidential elections ('niinistö'), Greece ('kreikka' or 'grekland') and Syria ('syyria' or 'syrien') in the political news topic for both languages. We see the rise and fall of the prominence of the terms according to their relevance in the news.

7 Conclusions and Future Work

In this paper we present a novel topic model, the *multilingual dynamic topic model* (ML-DTM), that combines dynamic topic modeling (DTM) and polylingual topic modeling (PLTM) to infer dynamic topics from aligned multilingual data. ML-DTM uses an extension of the DTM inference method of Bhadury et al. (2016) to aligned multilingual data.

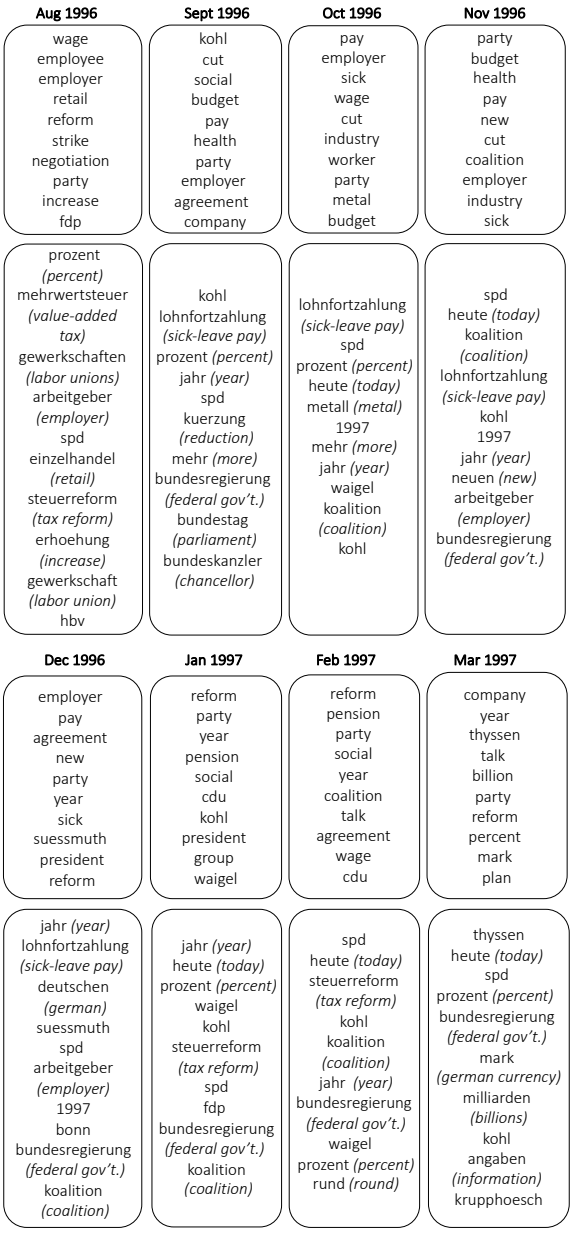

Figure 4: Top words of a topic concerning news about labor unions from the DE-NEWS dataset for English (top) and German (bottom) from Aug 1996 to March 1997. English translations of the German words excluding named entities are enclosed in parentheses.

We ran experiments on ML-DTM with parallel and comparable datasets. We compare cross-lingual topic alignment of PLTM and ML-DTM by evaluating the cosine similarities of topic vectors corresponding to named entity terms across languages for corresponding time slices. ML-DTM achieves similar performance to PLTM on DE-NEWS and the comparable dataset (YLE). We also demonstrate the ability of ML-DTM to detect

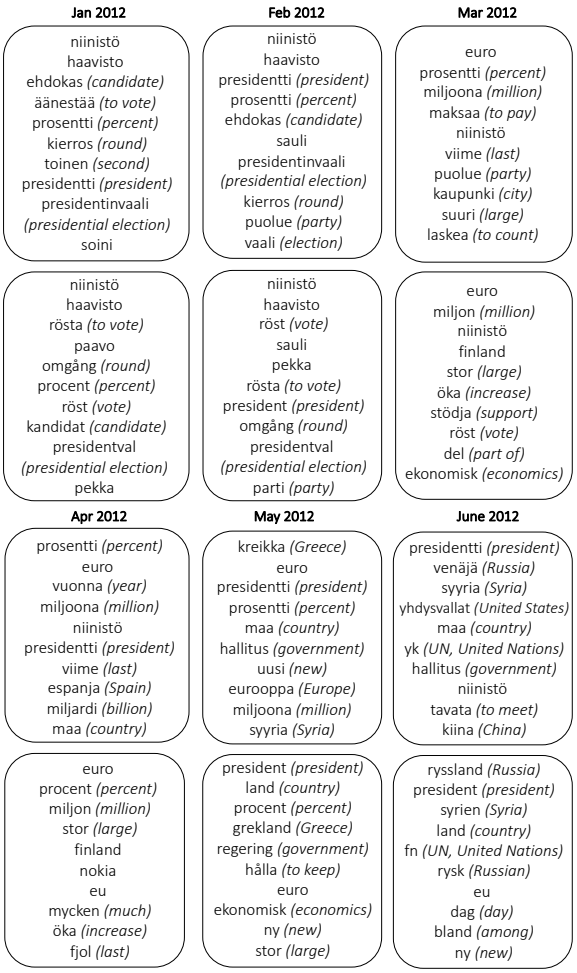

Figure 5: Top words of a topic on political news in Finland from the YLE dataset for Finnish (top) and Swedish (bottom) from Jan to June 2012. English translations of the words excluding named entities are enclosed in parentheses.

significant events regarding a topic through sudden changes in the prominent terms of the topic. This same method can also detect approximately when the event emerged and when it ended.

In a further experiment, we compared ML-DTM to the monolingual DTM, showing that ML-DTM achieves a consistently higher topic diversity within a single language.

We plan to run further experiments with ML-DTM using noisy datasets, such as historical news data where OCR errors might affect upstream tasks such as tokenization and lemmatization. We also plan to use named-entity recognition to improve our model such that named entities are treated as distinct items in the model's vocabulary, allowing us to track mentions of an entity across time slices and languages.

Historical news data covering a longer time

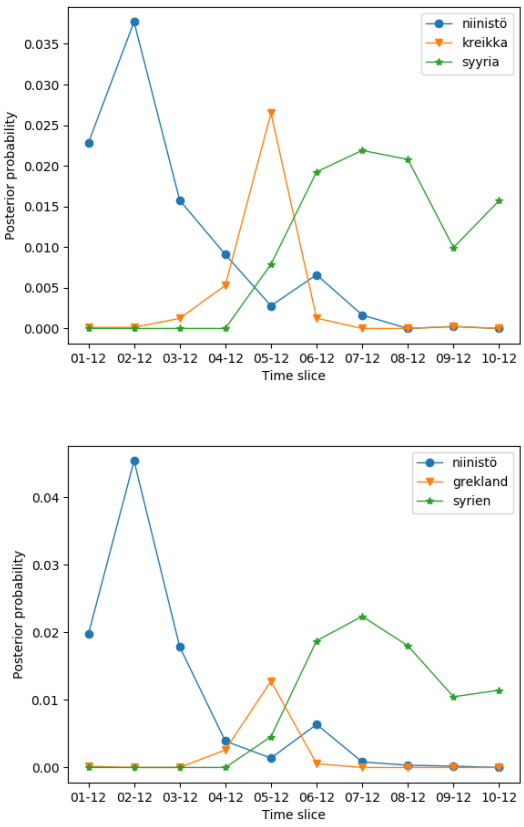

Figure 6: Posterior probabilities of salient terms in Finnish (top) and Swedish (bottom) related to events in the political news topic captured by ML-DTM from the YLE dataset.

span (several decades or more) would also enable us to study the changes in the use of words in a language and compare these changes with other languages. Historical news data from different regions would enable us to compare the way certain historical events were discussed in these places.

Acknowledgements

This work has been supported by the European Union's Horizon 2020 research and innovation programme under grants 770299 (NewsEye) and 825153 (EMBEDDIA).

References

Arnab Bhadury, Jianfei Chen, Jun Zhu, and Shixia Liu. 2016. Scaling up dynamic topic models. In *Proceedings of the 25th International Conference on World Wide Web*. International World Wide Web Conferences Steering Committee, pages 381–390.

David M Blei and John D Lafferty. 2006. Dynamic topic models. In *Proceedings of the 23rd interna-*

tional conference on Machine learning. ACM, pages 113–120.

David M Blei, Andrew Y Ng, and Michael I Jordan. 2003. Latent dirichlet allocation. *Journal of machine Learning research* 3(Jan):993–1022.

Jordan Boyd-Graber and David M Blei. 2009. Multilingual topic models for unaligned text. In *Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence*. AUAI Press, pages 75–82.

Lea Frermann and Mirella Lapata. 2016. A bayesian model of diachronic meaning change. *Transactions of the Association for Computational Linguistics* 4:31–45.

Thomas L Griffiths and Mark Steyvers. 2004. Finding scientific topics. *Proceedings of the National academy of Sciences* 101(suppl 1):5228–5235.

David Hall, Daniel Jurafsky, and Christopher D Manning. 2008. Studying the history of ideas using topic models. In *Proceedings of the conference on empirical methods in natural language processing*. Association for Computational Linguistics, pages 363–371.

Jagadeesh Jagarlamudi and Hal Daumé. 2010. Extracting multilingual topics from unaligned comparable corpora. In *European Conference on Information Retrieval*. Springer, pages 444–456.

Philipp Koehn. 2005. Europarl: A parallel corpus for statistical machine translation. In *MT summit*. volume 5, pages 79–86.

Eetu Mäkelä. 2016. Las: an integrated language analysis tool for multiple languages. *The Journal of Open Source Software* 1.

David Mimno, Hanna M Wallach, Jason Naradowsky, David A Smith, and Andrew McCallum. 2009. Polylingual topic models. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*. Association for Computational Linguistics, pages 880–889.

Masao Utiyama and Hitoshi Isahara. 2003. Reliable measures for aligning japanese-english news articles and sentences. In *Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1*. Association for Computational Linguistics, pages 72–79.

Jurgen Van Gael and Xiaojin Zhu. 2007. Correlation clustering for crosslingual link detection. In *IJCAI*. pages 1744–1749.

Thuy Vu, Ai Ti Aw, and Min Zhang. 2009. Feature-based method for document alignment in comparable news corpora. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, pages 843–851.

Ivan Vulić, Wim De Smet, Jie Tang, and Marie-Francine Moens. 2015. Probabilistic topic modeling in multilingual settings: An overview of its methodology and applications. *Information Processing & Management* 51(1):111–147.

Chong Wang, David Blei, and David Heckerman. 2008. Continuous time dynamic topic models. In *Proceedings of the Twenty-Fourth Conference on Uncertainty in Artificial Intelligence*. AUAI Press, Arlington, Virginia, United States, UAI'08, pages 579–586. http://dl.acm.org/citation.cfm?id=3023476.3023545.

Xuerui Wang and Andrew McCallum. 2006. Topics over time: a non-markov continuous-time model of topical trends. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, pages 424–433.

Xing Wei, Jimeng Sun, and Xuerui Wang. 2007. Dynamic mixture models for multiple time-series. In *Ijcai*. volume 7, pages 2909–2914.

Max Welling and Yee W Teh. 2011. Bayesian learning via Stochastic Gradient Langevin Dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*. pages 681–688.

Tze-I Yang, Andrew Torget, and Rada Mihalcea. 2011. Topic modeling on historical newspapers. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*. pages 96–104.